Molecular dynamics simulations of outer-membrane protease T from *E. coli* based on a hybrid coarse-grained/atomistic potential

2006 J. Phys.: Condens. Matter 18 S347

(http://iopscience.iop.org/0953-8984/18/14/S16)

INSTITUTE OF PHYSICS PUBLISHING JOURNAL OF PHYSICS: CONDENSED MATTER

J. Phys.: Condens. Matter **18** (2006) S347–S355 doi:10.1088/0953-8984/18/14/S16

Molecular dynamics simulations of outer-membrane protease T from *E. coli* based on a hybrid coarse-grained/atomistic potential

Marilisa Neri, Claudio Anselmi, Vincenzo Carnevale, Attilio V Vargiu and Paolo Carloni

International School for Advanced Studies (SISSA/ISAS) and INFM-DEMOCRITOS Modeling Center for Research in Atomistic Simulation, Via Beirut 4, I-34014 Trieste, Italy

Received 12 October 2005
Published 24 March 2006
Online at stacks.iop.org/JPhysCM/18/S347

Abstract

Outer-membrane proteases T (OmpT) are membrane enzymes used for defense by Gram-negative bacteria. Here we use hybrid molecular mechanics/coarse-grained simulations to investigate the role of large-scale motions of OmpT from *Escherichia coli* for its function. In this approach, the enzyme active site is treated at the all-atom level, whilst the rest of the protein is described at the coarse-grained level. Our calculations agree well with previously reported all-atom molecular dynamics simulations, suggesting that this approach is well suitable to investigate membrane proteins. In addition, our findings suggest that OmpT large-scale conformational fluctuations might play a role for its biological function, as found for another protease class, the aspartyl proteases.

(Some figures in this article are in colour only in the electronic version)

1. Introduction

Outer-membrane proteases T (OmpT) are membrane proteins expressed by Gram-negative bacteria such as *Yersinia pestis* and *Escherichia coli* [1–4]. They cleave antimicrobial peptides preferentially between two consecutive basic amino acids, activate human plasminogen, and degrade some recombinant heterologous proteins [5, 6].

The structure of OmpT from *Escherichia coli* has been solved at 2.6 Å resolution. The protein folds into a 10-strand antiparallel β-barrel conformation [7]. The extracellular part of the molecule contains a large negatively charged groove, which is consistent with the high specificity for positive substrates of the protein [2]. The deep groove is formed by loops L4 and L5 on one side and L1–L3 on the other (figure 1(a)).

The presence of a serine and a histidine (Ser^{99} and His^{212}) in the cleft led to the suggestion that OmpT was a novel-type serine protease [3]. Within this hypothesis, the scissile peptide bond should be attacked by the hydroxyl of the catalytic serine, which in serine protease is

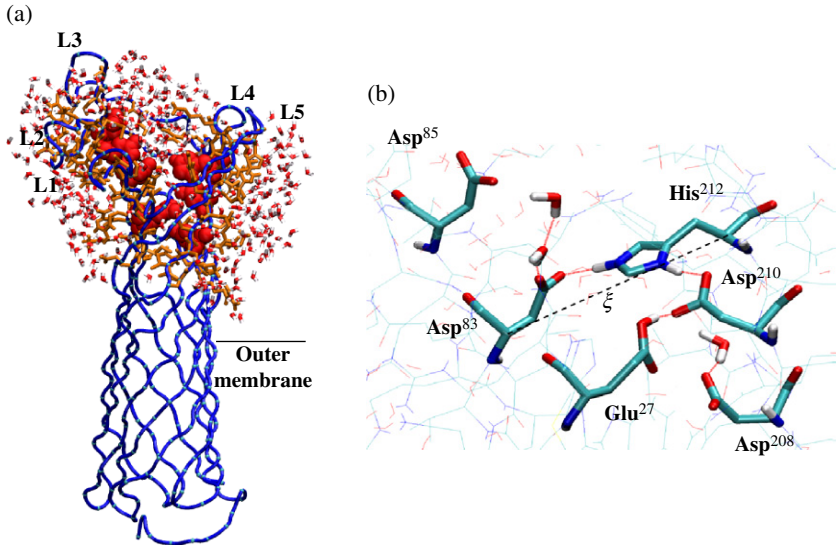

Figure 1. (a) Structure of OmpT from *Escherichia coli*. The MM, I and CG regions (see figure 2) are represented by red VdW sphere, orange licorice tubes and blue tubes, respectively. Water molecules are also shown. (b) OmpT active site: the residues in MM region are depicted with the licorice representation. H-bonds are represented as red lines. The thin lines represent groups included in the I region. The distance between C_α carbon of His^{212} and Asp^{83} (ξ) is represented as a dashed line.

usually activated by a histidine residue [8]. Then, an alternative mechanism of action was put forward, based on the following considerations: (i) the two catalytic His and Ser residues are much farther apart (∼9 Å) than in actual serine proteases; (ii) several Asp and Glu groups present in the cleft (figure 1(b)) seem to play a key role for catalysis, as shown by molecular biology experiments [3]. In this proposal, His^{212} and Asp^{83} groups activate a water molecule for a nucleophilic attack, while Asp^{83} and Asp^{85} contribute to polarize the substrate scissile peptide bond [4, 7, 9]. Molecular dynamics simulations (MD) on OmpT on a 10 ns timescale have further supported this hypothesis [9–11].

Here, we investigated the dynamics of OmpT on a relatively long timescale (70 ns) using a molecular mechanics/coarse-grained (MM/CG) scheme [12]. This approach provides an all-atom representation of the amino acid residues involved in the ligand binding region of an enzyme (i.e. the active site), whereas the remainder of the protein is treated with a widely used coarse-grained (CG) model (figure 2) [13–25]. Therefore, structural and dynamical properties of the active site can be investigated with relatively moderate computational resources.

In a previous work, our MM/CG approach turned out to reproduce well the structural and dynamical features of cytoplasmatic aspartyl proteases in complex with their substrates [12].

Here, OmpT key residues in the active site (Glu^{27}, Asp^{83}, Asp^{85}, Ser^{99}, Asp^{208}, Asp^{210} and His^{212}) are treated at the atomic detail (MM region). A shell of ∼6 Å constitutes the interface (I) between the MM residues and the CG region, which includes the remainder of the protein. Water molecules are treated explicitly in proximity of the MM and I regions (figures 1(a) and (b)).

The calculated structural properties of the proteins agree well with those obtained by all-atom MD simulations, suggesting that our MM/CG approach is valid also for treating membrane proteins. Furthermore, large-scale conformational fluctuations seem to play an

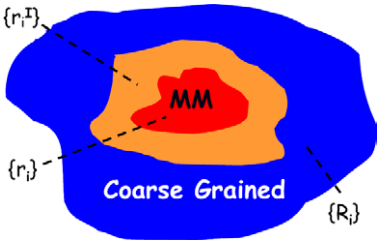

Figure 2. MM/CG Model: the MM (set of atoms $\{\mathbf{r}_i\}$), I (set of atoms $\{\mathbf{r}_i^{\mathrm{I}}\}$) and CG (set of atoms $\{\mathbf{R}_i\}$) regions are coloured in red, orange and blue, respectively.

important role, suggesting that some aspects of the mechanism of action of OmpT may be similar to those of other classes of enzymes [26–29].

2. Theory

2.1. The MM/CG model

A detailed description of the MM/CG approach is reported elsewhere [12]; therefore, we only summarize here the principles of the method. A small part of the protein (e.g. an enzymatic active site) is investigated with all-atom details, while the rest is treated with a CG approach by only considering C_α centroids. An interface region (I) is located between the two MM and CG regions, bridging the large discontinuity between full-atom and CG descriptions (figure 2). The total potential energy of the system reads

$$V = E_{\mathrm{MM}} + E_{\mathrm{CG}} + E_{\mathrm{I}} + E_{\mathrm{MM/I}} + E_{\mathrm{CG/I}} + E_{\mathrm{SD}},$$

where the first three terms represent the interactions within the MM, CG and I regions, respectively, whereas the fourth and fifth represent the cross-terms potentials. The last term, E_{SD}, mimics stochastic and frictional forces acting on the system, due to the solvent [30, 31].

E_{MM} is represented by the widely used GROMOS96 43a1 force field, with only polar hydrogens explicitly considered [32].

E_{CG} takes the following form:

$$E_{\mathrm{CG}} = \tfrac{1}{4}\sum_i K_{\mathrm{b}}(|\mathbf{R}_i - \mathbf{R}_{i+1}|^2 - b_{ii+1}^2)^2 + \sum_{i>j} V_0\left[1 - \exp\left(-B_{ij}(\left|\mathbf{R}_i - \mathbf{R}_j\right| - b_{ij})\right)\right]^2. \quad (1)$$

The first term in equation (1) takes into account bonded interactions between consecutive CG (C_α) centroids, identified by the position vectors $\mathbf{R}_i$ and $\mathbf{R}_{i+1}$, and K_{b} is the relative bond force constant[1]. b_{ij} is the equilibrium distance, corresponding to the native distance between CG atoms. The second term in equation (1) describes the non-bonded interactions between CG atoms. V_0 is the interaction well depth[2]. B_{ij} is the modulating exponent of the Morse potential.

In region I, all atoms are explicitly considered, as in the MM part, and consequently both E_{I} and $E_{\mathrm{MM/I}}$ energy terms have the same formulation of E_{MM} [12].

At the interface between the I and CG regions, bonds between consecutive C_α belonging to the I and CG regions ensure backbone connectivity [12]. In addition, a term describing the

[1] We chose $K_{\mathrm{b}} = 7.2 \times 10^6$ kJ mol^{-1} nm^{-4}, the typical force constant of a bond between sp^3 carbons in the GROMOS96 force-field [12, 32].

[2] The interaction well depth, V_0, was conveniently chosen equal to 5.3 kJ mol^{-1}. The non-bonded interactions are computed between CG atoms within a cut-off distance of $r_{\mathrm{cut}} = 10$ Å [12].

non-bonded interactions is added:

$$E_{\mathrm{CG/I}} = \frac{1}{2} \sum_{i \in [\mathrm{C}_\alpha, \mathrm{C}_\beta], j} V_0 \left[1 - \exp\left(-B_{ij}(\left|\mathbf{r}_i^{\mathrm{I}} - \mathbf{R}_j\right| - b_{ij})\right)\right]^2, \tag{2}$$

where the interface ith atom is either a C_α or a C_β atom (identified by the position vector $\mathbf{r}_i^{\mathrm{I}}$) and the factor $1/2$ stands for the interaction energy equally distributed between the two types of atoms. All the coefficients are chosen similarly to those in equation (1).

2.2. Calculation details

Our initial model is based on the crystal structure of S99A-G216K-K217 G OmpT from *E. coli*, solved at 2.6 Å resolution (PDB code 1I78, chain A) [7]. These mutations involve residues treated at the all-atom level in our computational scheme. Therefore, we restored the wild-type structure by reversing mutations by using the Swiss-Pdb Viewer program [33]. The protonation state of the residues present in the catalytic cleft was set as in [9]: Glu^{27} was protonated in ϵ, His^{212} was diprotonated, whilst all the Asp residues were taken as ionized (figure 1(b)). A water layer of 12 Å, centred on the MM region, was added, corresponding to 63 SPC water molecules [34]. As a result, the total system was composed of 2062 atoms.

The simulations were performed using a modified version of the Gromacs 3.2.1 program [31]. The leap-frog stochastic dynamics algorithm was used to integrate the equations of motion with a time step $\Delta t = 2$ fs and a friction coefficient $\gamma_i = m_i/\tau$, where $\tau = 0.5$ ps is the time constant for the coupling and m_i is the mass of ith particle. The temperature was maintained at 300 K using a Berendsen thermostat [35]. The SHAKE algorithm [36] was used to keep the distance of bonds containing hydrogens fixed. No cut-off was used for the non-bonded interactions.

The system was first relaxed with a 100 ps run with positional restraints on solute positions and then simulated for 70 ns without positional restraints.

The property on which this work is focused on is the large-scale conformational fluctuations. Specifically, we calculate concerted motions, which are related to the degree of correlation of pairs of residues at equal times. This information is summarized in the covariance matrix, $\mathcal{C}$, whose elements are given by

$$\mathcal{C}_{ij,\mu\nu} \equiv \langle x_{i,\mu} x_{j,\nu} \rangle = \sum_l \mathbf{v}_i^l \mathbf{v}_j^l \lambda_l, \tag{3}$$

where the average $\langle . \rangle$ denotes the usual canonical thermodynamics average, $x_{i,\mu}$ is the deviation of ith C_α along the μ axis (μ and ν denote the Cartesian components, x, y and z), the vector $\mathbf{v}^l$ and the scalar λ_l are the lth eigenvector and the lth eigenvalue of matrix $\mathcal{C}$, respectively, ordered in such a way that $\lambda_l > \lambda_{l+1}$. This matrix contains the full three-dimensional information about pair correlations and its linear size is $3N$, with N number of residues. Typically it is important to quantify the relative degree of correlation of any pair of residues; in this case one can consider the normalized reduced covariant matrix (of linear size N), which is defined as

$$\mathcal{C}_{ij} = \frac{\mathcal{C}_{ij,\nu\nu}}{\sqrt{\mathcal{C}_{ii,\nu\nu}\mathcal{C}_{jj,\mu\mu}}}. \tag{4}$$

3. Results

We compared our results with those of all-atom molecular dynamics simulation. Then, we investigated the motion of the protein frame during the MM/CG simulation. Finally, we analysed the structural properties of the important residues for the protein function.

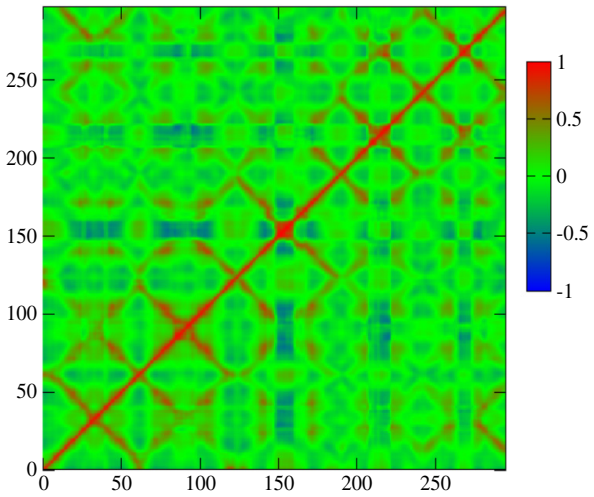

Figure 3. Normalized reduced covariance matrix of OmpT C_α atoms.

3.1. Comparison with previous MD calculations

The covariance matrix, $\mathcal{C}$, (figure 3) describes the degree of correlation of pairs of residues at the same time (see method section). The accuracy of our method for reproducing the all-atom MD on the protein frame has been established by a comparison with $\mathcal{C}$-derived properties obtained with a 10 ns MD simulation on the same system [10].

(i) The root mean square fluctuations (RMSF), given by the square root of the trace of the covariance matrix in equation (3), were well reproduced, with a correlation coefficient between the two sets of data of 0.79 (figure 4(a)). We also compared the RMSF given by the crystallographic structure with those obtained by using the MM/CG simulation, finding a correlation coefficient of ∼0.66.

(ii) The projections of the MM/CG eigenvectors associated with the five largest eigenvalues of $\mathcal{C}$ (which represent the large-scale motion of the system) onto the corresponding eigenvectors calculated with MD overlapped by 75%.

(iii) Finally, the eigenvalues of $\mathcal{C}$ calculated with the two potentials were fitted with a exponent of 0.97 (figure 4(b)), showing that the two entries practically coincide. These values are proportional to the characteristic timescale of the collective vibrational excitations. Therefore, the rate of molecular events in MM/CG simulations practically corresponds to that of MD simulations.

In addition, the stability of the protein with respect to the x-ray structure [7] has been shown by the low value of the RMSD of the C_α carbons during the dynamics (figure 5).

3.2. Large-scale motions of the protein

Based on the calculation of $\mathcal{C}$ (figure 3), we suggested that the motion of the groups ranging from 150 to 160 (constituting the loop L3 surrounding the active site, see figure 1(a)) is anticorrelated with that of the scaffold of the protein, in particular with groups 20–50 (belonging to the β-sheets $\beta1$ and $\beta2$ and the loop L1 bridging them), 70–110 ($\beta3$, L4 and $\beta4$), 170–180 ($\beta6$), 200–210 ($\beta7$), 220–230 ($\beta8$), 250–265 ($\beta9$), 272–285 ($\beta10$). In addition, the β-sheets from $\beta1$ to $\beta10$ are strongly correlated among them. As a result, L3 moves in a counter direction

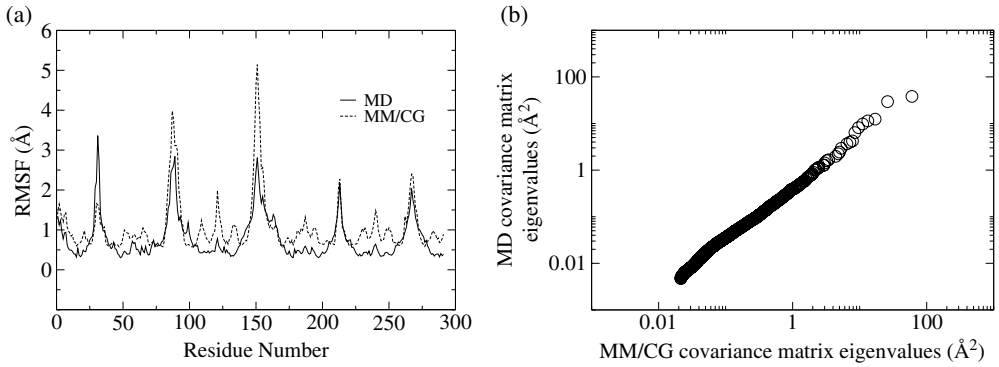

Figure 4. (a) Root mean square fluctuations of the various residues in OmpT obtained from MD and MM/CG simulations. (b) Scatter plot of equally ranking eigenvalues of the MD and MM/CG covariance matrices.

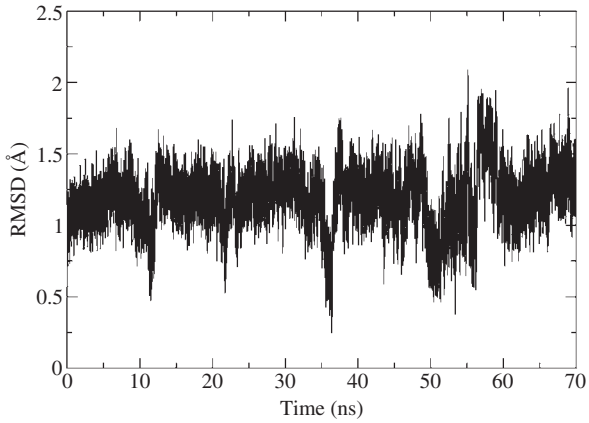

Figure 5. RMSD computed on the C_α atoms of the MM residues.

with respect to the compact and rigid β-barrel core (and the bridging loops L1, L2, L4 and L5). This is indicated more clearly in figure 6, showing the four most significant eigenvectors of the covariance matrix, that account for ∼50% of the overall residue mobility. These motions might be essential for driving the residues in the active site to a specific configuration and, therefore, could be important for OmpT catalytic activity.

3.3. H-bond network in the active site

All the residues putatively involved in the enzymatic hydrolysis [4] interact with each other. During the dynamics, Asp^{210} keeps its original position by H-bonding to His^{212} and by forming a direct or water-mediated H-bond with Glu^{27}. At the same time, Glu^{27} interacts through a water bridge or a direct H-bond with Asp^{208}. Asp^{83} and Asp^{85} interact through (generally) two or three bridging waters. Table 1 shows that most of these structural features reproduce the x-ray data.

An important feature of the dynamics in the active site concerns the distance between the key residues His^{212} and Asp^{83} (ξ), which also represents the width of the cleft (figure 7). During most of the dynamics, ξ oscillates around an average value of 9.7 Å corresponding to

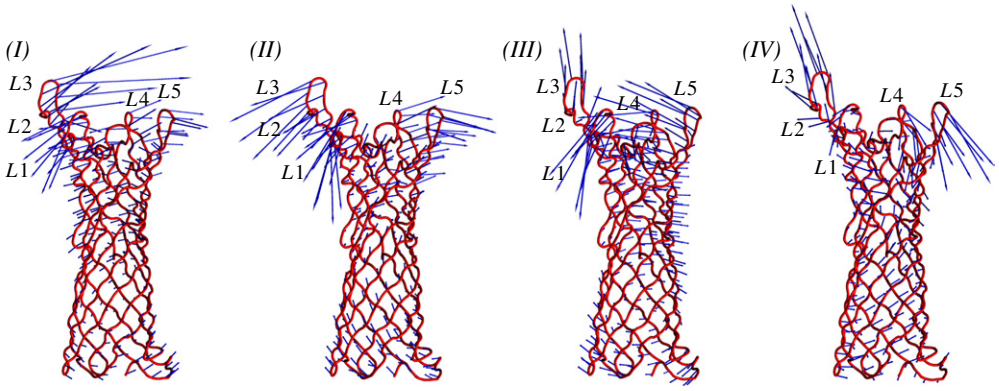

Figure 6. The eigenvectors relative to the largest four eigenvalues are represented (for each of the atomic components) as arrows of width and length proportional to the modulus. The protein is shown in trace representation.

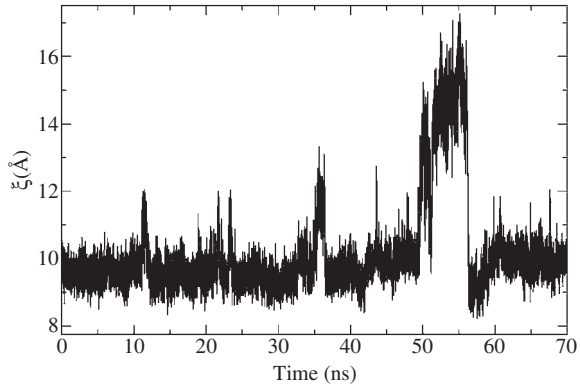

Figure 7. Distance between the C_α carbons of His^{212} and Asp^{83}, ξ, as a function of time.

Table 1. C_α–C_α distances of pairs of residues in the active site calculated from the MM/CG simulation and the crystallographic structure. MM/CG values were estimated by averaging the distances over the whole simulation, whereas the errors were estimated by the corresponding standard deviations.

	C_α–C_α MM/CG distance (Å)	C_α–C_α crystallographic distance (Å)
His^{212}–Asp^{83}	10.2 ± 1.5	12.5
His^{212}–Asp^{210}	6.2 ± 0.2	6.5
Asp^{83}–Asp^{85}	6.7 ± 0.4	6.9
Glu^{27}–Asp^{208}	8.4 ± 0.7	8.8
Glu^{27}–Asp^{210}	9.1 ± 0.9	9.2

Asp^{83} and His^{212} forming a direct H-bonding interaction. However, at times, ξ increases for a few nanoseconds, allowing a water molecule to form a bridge between the two residues. In some cases (e.g. at ∼50 ns) more than one water molecule (two or three) is inserted between Asp^{83} and His^{212} for a short time.

4. Discussion and conclusion

Our MM/CG calculations on OmpT were able to reproduce both local and mesoscopic properties obtained with MD simulations. As the latter are reproduced by means of a CG representation, this finding points to the suitability of CG approaches not only for globular proteins, which are immersed in a homogeneous medium (i.e. water) [12], but also for membrane proteins, whose environment features discontinuities at the water–polar head and apolar chain–polar head interfaces.

Our calculations suggest that the distance between the putative catalytic dyad His^{212} and Asp^{83} [4] periodically (with a 12 ns period) increases from ∼10 to 15 Å. This opening motion origins from the large-scale fluctuations of L2, which are correlated to those of loop L4 (figure 6). In fact, His^{212} is located in $\beta 8$, close to the loop L4, whereas Asp^{83} is located in $\beta 3$, close to the loop L2, at the other side of the catalytic cleft. As a result, a water molecule can be accommodated in a suitable position for the enzymatic hydrolysis. At the speculative level, this motion might therefore play a role for catalysis. This mechanism shares some resemblance with those proposed for other proteases (i.e. the aspartyl proteases HIV-1 PR [37] and BACE [38]), where the concerted motion of the loops is crucial to drive the substrate in an active conformation [39, 29, 40]. Simulations of the complex with OmpT substrate are required to assess this issue.

Limitations of the model. Two main limitations of the model used here are: (i) the presence of a repulsive potential at the boundaries of the water sphere used to avoid diffusion; (ii) the lack of a membrane environment. Work is in progress to treat hydration more accurately by performing MD simulations at constant chemical potential and to include lipid molecules treated explicitly with a CG model [41, 42].

Acknowledgments

The authors are indebted to José Faraldo-Gómez (Weill Medical College, NY, USA) for providing the data about OmpT 10 ns MD simulation in a hydrated lipid bilayer. We also thank Amos Maritan (Università di Padova, Italy), Michele Cascella (EPFL, Switzerland), Gianluca Lattanzi (Università di Bari, Italy) and Marco Berrera (SISSA/ISAS, Italy) for suggestions and useful discussions. This work was supported by MURST-COFIN.

References

[1] Mangel W F, Toledo D L, Brown M T, Worzalla K, Lee M and Dunn J J 1994 *Methods Enzymol.* **244** 384–99
[2] Dekker N, Cox R C, Kramer R A and Egmond M R 2001 *Biochemistry* **40** 1694–701
[3] Kramer R A, Dekker N and Egmond M R 2000 *FEBS Lett.* **468** 220–4
[4] Kramer R A, Vandeputte-Rutten L, de Roon G J, Gros P, Dekker N and Egmond M R 2001 *FEBS Lett.* **505** 426–30
[5] Guina T, Yi E C, Wang H, Hackett M and Miller S I 2000 *J. Bacteriol.* **182** 4077–86
[6] Stumpe S, Schmid R, Stephens D L, Georgiou G and Bakker E P 1998 *J. Bacteriol.* **180** 4002–6
[7] Vandeputte-Rutten L, Kramer R A, Kroon J, Dekker N, Egmond M R and Gros P 2001 *EMBO J.* **20** 5033–9
[8] Barret A J, Rawlings N D and Woessner J F 1998 *Handbook of Proteolytic Enzymes* (San Diego, CA: Academic)
[9] Baaden M and Sansom M S P 2004 *Biophys. J.* **87** 2942–53
[10] Faraldo-Gomez J D, Forrest L R, Baaden M, Bond P J, Domene C, Patargias G, Cuthbertson J and Sansom M S 2004 *Proteins* **57** 783–91
[11] Bond P J and Sansom M S 2004 *Mol. Membr. Biol.* **21** 151–61
[12] Neri M, Anselmi C, Cascella M, Maritan A and Carloni P 2005 *Phys. Rev. Lett.* at press
[13] Brooks B and Karplus M 1985 *Proc. Natl Acad. Sci. USA* **82** 4995–9
[14] McCammon J A, Gelin B R, Karplus M and Wolynes P G 1976 *Nature* **262** 325–6
[15] Noguti T and Go N 1982 *Nature* **296** 776–8

[16] Swaminathan S, Ichiye T, van Gunsteren W and Karplus M 1982 *Biochemistry* **21** 5230–41
[17] Tirion M M 1996 *Phys. Rev. Lett.* **77** 1905–8
[18] Bahar I, Atilgan A R, Jernigan R L and Erman B 1997 *Proteins* **29** 172–85
[19] Hinsen K 1998 *Proteins* **33** 417–29
[20] Atilgan A R, Durell S R, Jernigan R L, Demirel M C, Keskin O and Bahar I 2001 *Biophys. J.* **80** 505–15
[21] Doruker P, Jernigan R L and Bahar I 2002 *J. Comput. Chem.* **23** 119–27
[22] Nielsen S O, Lopez C F, Srinivas G and Klein M L 2004 *J. Phys.: Condens. Matter* **16** R481–512
[23] Navizet I, Cailliez F and Lavery R 2004 *Biophys. J.* **87** 1426–35
[24] Micheletti C, Carloni P and Maritan A 2004 *Proteins* **55** 635–45
[25] Neri M, Cascella M and Micheletti C 2005 *J. Phys.: Condens. Matter* **17** S1581–93
[26] Piana S, Carloni P and Parrinello M 2002 *J. Mol. Biol.* **319** 567–83
[27] Yang H *et al* 2003 *Science* **302** 262–6
[28] Rod T H *et al* 2003 *Proc. Natl Acad. Sci. USA* **100** 6980–5
[29] Perryman A L, Lin J H and McCammon J A 2004 *Protein Sci.* **13** 1108–23
[30] Doi M 1996 *Introduction To Polymer Physics* 1st edn (Great Britain: Oxford Science)
[31] van der Spoel D, Lindahl E, Hess B, van Buuren A R, Apol E, Meulenhoff P J, Tieleman D P, Sijbers A L T M, Feenstra K A, van Drunen R and Berendsen H J C 2004 *Gromacs User Manual Version 3.2* http://www.gromacs.org/documentation/paper_manuals.php
[32] van Gunsteren W F, Billeter S R, Eising A A, Hünenberg P H, Krüger P, Mark A E, Scott W R P and Tironi I G 1996 *Biomolecular Simulation: The GROMOS96 Manual and User Guide* (Zurich: Hochschulverlag AG an der ETH Zurich)
[33] Guex N and Peitsch M C 1997 *Electrophoresis* **18** 2714–23
[34] Berendsen H J C, Postma J P M, van Gunsteren W F and Hermans J 1981 *Intermolecular Forces* ed B Pullman (Dordrecht: Reidel) pp 331–42
[35] Berendsen H J C, Postma J P M, van Gunsteren W F, DiNola A and Haak J R 1984 *J. Chem. Phys.* **81** 3684–90
[36] Ryckaert J P, Ciccotti G and Berendsen H J C 1977 *J. Comput. Phys.* **23** 327–41
[37] Miller M, Schneider J, Sathyanarayana B K, Toth M V, Marshall G R, Clawson L, Selk L M, Kent S B H and Wlodawer A 1989 *Science* **246** 1149–52
[38] Hong L, Koelsch G, Lin X, Wu S, Terzyan S, Ghosh A K, Zhang X C and Tang J 2000 *Science* **290** 150–3
[39] Piana S, Carloni P and Parrinello M 2002 *J. Mol. Biol.* **319** 567–83
[40] Cascella M, Micheletti C, Rothlisberger U and Carloni P 2005 *J. Am. Chem. Soc.* **127** 3734–42
[41] Shelley J C, Shelley M Y, Reeder R C, Bandyopadhyay S and Klein M L 2001 *J. Phys. Chem.* B **105** 4464–70
[42] Marrink S J, de Vries A H and Mark A E 2004 *J. Phys. Chem.* B **108** 750–60